

Uso de fuentes de información geográfica voluntarias en proyectos de ingeniería

Cruz E. Borges⁽¹⁾, Ander Pijoan^{(1)}, Gorka Sorrosal⁽¹⁾,
Iraia Oribe-Garcia⁽¹⁾, Mikel González⁽¹⁾ y Oihane Kamara Esteban⁽¹⁾*

- (1) Deusto Institute of Technology - DeustoTech Energy,
Universidad de Deusto, Avda. Universidades 24,
48007, Bilbao, Euskadi, España.
{cruz.borges,ander.pijoan,gsorrosal}@deusto.es
{iraia.oriobe,mikel.astorga,oihane.esteban}@deusto.es

Resumen

Conseguir fuentes de datos geoespaciales fiables y accesibles es, en muchos de los casos, la tarea más complicada y costosa de un proyecto de ingeniería. Aún así, uniéndose a la reciente corriente *open data*, muchas autoridades públicas han empezado a liberar sus fuentes de datos SIG. Por otro lado existen comunidades de usuarios que han generado bases de datos SIG mediante contribuciones voluntarias como por ejemplo OpenStreetMap (OSM).

Ambos casos, sin embargo, presentan varios problemas a resolver tales como la existencia de información en ficheros no interoperables (PDF es el caso más típico) o en distintos estándares. Por otro lado, a pesar de que la información geográfica creada por la comunidad de usuarios puede carecer de un amplio nivel de detalle en algunas zonas, sus datos siguen estrictos estándares de codificación, además de proveer de excelentes herramientas para su manejo y visualización.

Con el fin de sortear estos obstáculos, proponemos usar la siguiente metodología: a) evaluar si los datos geográficos actualmente proporcionados por la comunidad son suficientes para cumplir el objetivo del proyecto b) utilizar (o incluso crear) herramientas para transformar estos datos al formato de datos OSM (y añadirlos al repositorio correspondiente en la medida de lo posible) c) por último, usar las herramientas existentes para acceder, editar, visualizar o analizar estos datos. De esta forma, obtenemos lo mejor de los dos mundos: un conjunto de datos completo y buenas herramientas para su manipulación.

Para finalizar, presentaremos varios casos de estudio en los que se ha aplicado satisfactoriamente esta metodología (tanto por nosotros como por otros grupos): crecimiento vegetativo de la demanda eléctrica, inspección de líneas eléctricas mediante vehículos aéreos no tripulados, indicadores de calidad medioambiental, Technical Cognitive Assistance Systems, Election Results Maps, Flight Gears Terrain Simulator o el Proyecto FireFIGHT.

Palabras Clave: Open Data, Software Libre, OpenStreetMap

Abstract

Obtaining reliable and accessible GIS data is often the most difficult and costly part of an engineering project. Following new tendencies, public authorities have started to create GIS repositories or, in some cases, communities have collected and made the data publicly available in initiatives such as OpenStreetMap (OSM). However, in both cases there are some challenges that must be faced. For example, Open Data repositories are usually encoded in several standards and/or in a non computer friendly format (e.g. PDF format). In addition, community-made data may present gaps but generally are encoded following standards and provide very good tools to edit and visualise it.

In order to solve these two problems we propose here the use of the following three steps methodology: *a)* evaluate if the data actually provided by the community is sufficient *b)* use (or create) the tools to transform these data to the OSM format (and upload it if possible) *c)* and finally, use their tools to access, edit, visualise or analyse the data. In this way, we have the best of the two worlds: a complete dataset and good tools to manipulate it.

Moreover, we will present here three engineering problems where we have successfully applied this methodology: Long-Term Load Forecast, Overhead Power Line inspection with UAV, Environmental quality index. Finally we will present other successful projects that have used OSM data and tools: Technical Cognitive Assistance Systems, Election Results Maps, Flight Gears Terrain Simulator and FireFIGHT.

Key words: Open Data, Free Software, OpenStreetMap

1. Introducción

Multitud de proyectos de ingeniería requieren información geoespacial, bien para su realización o bien como información complementaria en el modelado. Sin embargo, la complejidad intrínseca y la escasa accesibilidad a dichos datos han hecho que el uso de información geoespacial sea anecdótico. Por otro lado, el desarrollo de dispositivos de medidas GPS y el acceso masivo a Internet está cambiando el panorama hasta el punto de hablarse de la «revolución GIS».

Los dispositivos GPS han democratizado el acceso a la toma de medidas topográficas permitiendo que la ciudadanía se involucre en la realización de mapas. Más aún, el auge de los dispositivos de navegación (incluidos también en los *smartphones*) ha despertado la conciencia social sobre la cartografía, retroalimentando, a su vez, tanto el interés en los propios mapas como el desarrollo de aplicaciones que los usen.

Todo ello, unido a la mayor demanda de transparencia por parte de la ciudadanía, ha creado el caldo de cultivo propicio para la aparición de movimientos sociales que exigen la apertura de todas las bases de datos financiadas con fondos públicos, especialmente

las basadas en datos geoespaciales, al ser estas últimas excepcionalmente caras y de difícil obtención por otros medios. Es más, estos movimientos sociales no se han limitado a ser actores pasivos, sino que, mientras los gobiernos toman decisiones sobre el tema, han comenzado a desarrollar sus propias bases de datos, de libre acceso, denominadas comúnmente como «fuentes de información geográfica voluntarias». Así mismo también han desarrollado multitud de herramientas de obtención, manipulación, edición y visualización de dichos datos (obviamente desarrolladas como software libre), creando todo un nuevo ecosistema alrededor del movimiento.

Sin embargo, no todo son buenas noticias. Estas bases de datos, al depender exclusivamente de los datos aportados por sus usuarios, pueden presentar heterogeneidades en el grado de cobertura de sus datos, sobre todo en zonas con baja densidad de población. Mientras que en las ciudades, el grado de detalle raya lo absurdo (se tienen geolocalizadas hasta las papeleras, carteles de información turística o las farolas), en localidades rurales es posible que ni siquiera las vías estén disponibles. Ahora bien, de existir datos, la calidad de estos es alta tanto en exactitud como en actualización [25].

Recientemente, muchas administraciones públicas han comenzado a facilitar el acceso a sus bases de datos geográficas (como por ejemplo Catastro [14] o GeoData Euskadi [39]). Estas bases de datos son más homogéneas aunque pueden no estar tan actualizadas, presentar problemas de codificación o tener la información dispersa en cientos de ficheros. Es más, los datos suelen darse en crudo y es preciso usar herramientas GIS caras y complejas que para poder ser modificados o visualizados. No obstante, este no es el caso en la principal fuente de datos geográficos voluntarios, OpenStreetMap (OSM) [33], ya que la comunidad de colaboradores y desarrolladores de OSM ha generado tanto una base de datos de información geográfica muy completa, como las herramientas para editarla [29], visualizarla [30] o analizarla [28] de forma fácil y amigable.

Desde un punto de vista del desarrollo, es claro que las bases de datos de información geográfica voluntaria (principalmente OSM) y las herramientas adyacentes presentan las mismas ventajas que presenta el software libre en general, aunque también sus mismos defectos. Esto es, la información disponible relativa al lugar donde se va a realizar un determinado proyecto puede que no tenga suficiente grado de detalle aunque por otro lado se dan todas las facilidades para subsanar este problema. Es en estos casos cuando las fuentes de datos gubernamentales pueden ser de gran ayuda si se realiza una importación masiva de información con el fin de poder utilizar las herramientas propias de OSM. De esta forma, se evita tanto el mantenimiento de la base de datos geoespacial como se realiza una acción beneficiosa para toda la comunidad.

Nótese que no siempre es posible realizar este último paso pues los datos necesarios en el proyecto pueden tener restricciones de licencia o ser incompatibles con la política de datos usada por la Fundación OSM [31]. Aún en este caso, es muy simple replicar parte (o toda) la infraestructura de OSM en un servidor local y de esta forma añadir la ventaja de tener el control absoluto sobre esta (aunque aumente el coste de mantenimiento).

Por todo ello, en las siguientes sección presentamos tres proyectos en los que ha

sido clave el uso de las herramientas o datos proporcionados por OSM para el correcto desarrollo de los mismos.

2. Ejemplos de proyectos que usan fuentes de información geográfica voluntaria

2.1. Crecimiento vegetativo de la demanda eléctrica

El objetivo de este proyecto es construir un modelo de predicción de la demanda de energía eléctrica a largo plazo que tenga en cuenta no solo el «cuándo» sino el «dónde». Este modelo permitirá a las compañías eléctricas predecir en qué momento qué infraestructuras pueden sobrecargarse y por lo tanto poder actuar con la suficiente antelación.

El principal requisito del proyecto fue la necesidad de que las fuentes de datos fueran lo más accesibles posible, al menos en el territorio español. Esto es, todos los datos de entrada tenían que estar codificados mediante estándares y estar en formato digital. En base a este condicionante, el modelo concebido se compone de dos fases: predicción global y reparto local.

2.1.1. Predicción Global

El crecimiento de la demanda de potencia eléctrica en una zona depende directamente de dos factores: cuántos clientes hay y cuánto consume cada cliente. Al incremento en el número de clientes se le denomina, tradicionalmente, «crecimiento horizontal» (pues requiere ocupar nueva superficie). Por otro lado, al aumento en el consumo de cada cliente se le conoce tradicionalmente como «crecimiento vertical».

Obviamente, cada uno de estos apartados necesita de mecanismos de predicción diferenciados pues su naturaleza es muy distinta. En esta parte del modelo, se realiza una predicción global de los parámetros antes mencionados, entendiéndose por global, el desarrollo de una predicción acerca de la cantidad y tipo de clientes que van a aparecer así como del aumento en el consumo de los clientes ya existentes.

Con respecto al crecimiento vertical, es conocido que la demanda de potencia eléctrica depende fundamentalmente de la temperatura y de ciertos índices macroeconómicos [24]. Es por ello que en este proyecto ajustamos un modelo lineal a la serie temporal anual de la máxima demanda de potencia eléctrica. En este modelo usamos como variables predictoras tanto las temperaturas máxima y mínima anuales como la población activa por sectores.

La información acerca de la potencia máxima anual está (obviamente) en posesión de la compañía eléctrica, las temperaturas máximas y mínimas podían¹ ser obtenidas con

¹Ahora es necesario pagar por ellos o pertenecer a una de las entidades bonificadas.

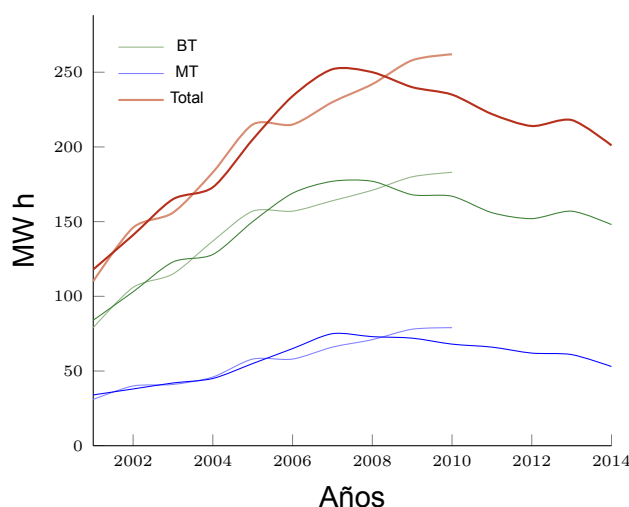


Figura 1: Resultados de la predicción global del crecimiento vertical.

facilidad de la Agencia Estatal de Meteorología (AEMET) [23] y los datos de la población activa pueden ser recogidos del Instituto Nacional de Estadística (INE) [9]. Es necesario que estos datos estén disponibles desagregados en la zona a estudiar, que en general coincide con la provincia. Tanto los datos de la AEMET como del INE están segmentados a ese nivel. Por último, para realizar la predicción de la demanda eléctrica, es necesario contar con una estimación tanto de las temperaturas como de la población activa. La AEMET ofrece predicciones a largo plazo para las temperaturas (con modelos de cambio climático) mientras que para la población activa es necesario analizar distintos escenarios macro-económicos. Esto puede presentar un problema pues la generación de estos escenarios de forma realista dista de ser trivial. Aún así, las grandes compañías eléctricas suelen contar internamente con este tipo de predicciones para múltiples fines. Los resultados tanto del ajuste como de la predicción del modelo se encuentra en la figura 1.

En cuanto a la predicción del número de nuevas viviendas creadas, los modelos encontrados en la literatura estudian otros factores como, por ejemplo: la dispersión [37], la relación entre el número de habitantes y el número de ciudades de la provincia [15] o el precio [36], pero no el número de nuevas viviendas.

Aunque es obvio que esta magnitud tiene relación con algún índice de variables macro-económicas, no ha sido posible encontrar ninguna publicación en este sentido, por lo que hemos decidido construir un modelo ARIMA [5] mucho más sencillo ya que solo depende de la serie histórica de viviendas construidas.

A pesar de que el dato anterior está recogido en las estadísticas del INE, en este caso en concreto, hemos decidido estimar el número de viviendas construidas mediante el número de nuevos suministros que han permanecido activos en la compañía eléctrica al menos durante ese año. De nuevo, este valor está en posesión de la compañía eléctrica que quiera utilizar este modelo y es posible que sea más fiel para el objetivo final del análisis. La figura 2 recoge el ajuste del modelo y su predicción.

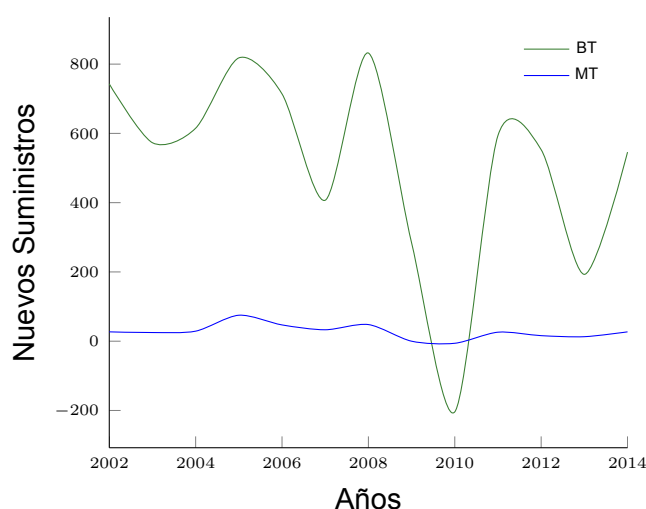


Figura 2: Resultados de la predicción global del crecimiento horizontal.

2.1.2. Reparto local

Una vez hecha la predicción global, es necesario repartirla a lo largo del territorio de estudio. Para ello, proponemos dos modelos: el modelo *bitmap* y el modelo vectorial.

Modelo *Bitmap*: En este modelo dividimos la zona de estudio en celdas de tamaño uniforme. A continuación, se identifican las zonas en las que es posible realizar nuevas construcciones y se calcula el número máximo de suministros que es capaz de aceptar dicha zona. El siguiente paso consiste en asignar una puntuación a cada una de estas zonas basándose en las distancias a distintas infraestructuras y el grado de ocupación de la celda. Posteriormente, se transforman estas puntuaciones en votos para luego ser asignados los nuevos suministros predichos por el modelo global siguiendo la clásica Ley D'Hont. Finalmente, se asignan los nuevos suministros a la infraestructura eléctrica más cercana que no esté saturada y esté a menos de 1,5 Km.

Cada fase del algoritmo necesita disponer de información de distinta índole. En primer lugar, es necesario localizar las zonas de posible construcción. Los Planes de Ordenación del Territorio serían el lugar natural al que acudir en estos casos, pero desgraciadamente estos planes se realizan a nivel municipal y cada municipio usa un mecanismo distinto, muchos de ellos ni siquiera digitales. Otra fuente que provee este tipo de información es el Catastro.

Actualmente, la Dirección General de Catastro ha liberado parte de la información que posee en sus bases de datos, incluida información acerca de las parcelas en las que es posible edificar. El problema reside en que el formato en el que se encuentran dichos datos no es muy amigable (un ShapeFile con la información geométrica junto con un archivo CSV con la información catastral). Por ello, se decidió construir una herramienta *ad hoc* que fusionase ambos conjuntos de datos y los transformase en un fichero susceptible de



Figura 3: Densidad de población interpolada mediante SVM.

ser importado a la base de datos de OSM [1]. Nótese que en algunas ciudades no sería necesario realizar este paso, pues es posible que la cobertura de OSM de la ciudad sea incluso mejor.

Una vez localizadas las parcelas urbanizables, se les asigna una capacidad máxima en función de su superficie y de la densidad de suministros que contenga la celda. A fin de estimar la densidad en las celdas sin suministros, se ajusta un modelo no lineal mediante una Support Vector Machine (SVM) [38] (entrenado con los valores de todas las celdas) y se usa el valor dado por el modelo como el máximo. La figura 3 presenta el resultado del ajuste de dicho modelo.

Finalmente, es necesario poseer una lista con las infraestructuras relevantes para evaluar la calidad de las celdas, su peso relativo y, obviamente, su localización. Para los dos primeros puntos, hemos realizado una encuesta muy simple al claustro de la Universidad de Deusto. El listado de infraestructuras relevantes, así como la distribución de pesos que se le asignó a cada infraestructura pueden ser consultados en la figura 4. Nótese que para el modelo final, se ha escogido la mediana como valor representativo de cada infraestructura normalizando los valores entre -1 y 1. Para el tercer punto, hemos vuelto a utilizar los datos de OSM localizando las infraestructuras y filtrando sus geometrías usando el diccionario del cuadro 1.

El algoritmo se ejecuta y genera la predicción del crecimiento de la demanda para el año siguiente. Usando dicho dato como real, podemos iterar y predecir los años necesarios; de esta forma se consiguen los valores de potencia demandada totales que se tienen en cada celda al sumar las demandas de todos los clientes (nuevos y ya estableci-

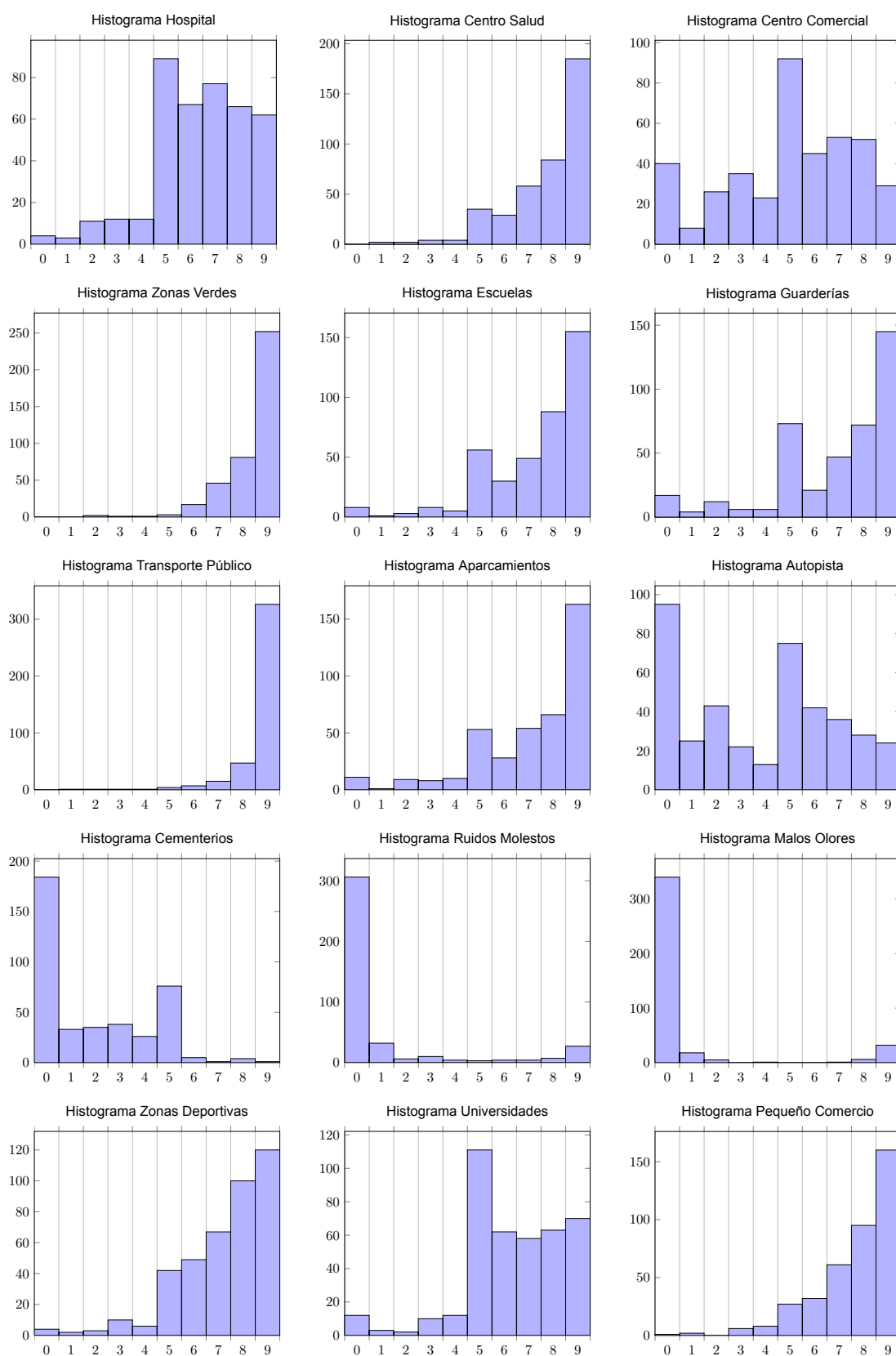


Figura 4: Resultados de la encuesta.

Infraestructura	Clave=Valor
Hospital	<i>amenity=hospital</i>
Centro de Salud	<i>amenity=doctors</i>
Autopista	<i>highway=motorway</i>
Transporte Público	<i>public_transport=station</i>
Parking	<i>amenity=parking</i>
Centro Comercial	<i>shop=mall</i>
Zona Verde	<i>leisure=park</i>
Polideportivo	<i>leisure=sports_centre</i>
Universidad	<i>amenity=university</i>
Escuela	<i>amenity=school</i>
Guardería	<i>amenity=kindergarten</i>
Cementerio	<i>landuse=cemetery</i>
Zona de Ruido	<i>amenity=nightclub</i>
Hedor	<i>landuse=landfill</i>
Pequeño Negocio	<i>landuse=retail</i>

Cuadro 1: Tabla de equivalencias de infraestructuras a tags OSM.

dos) así como su agregación en las infraestructuras eléctricas que les corresponden. La figura 5 presenta un ejemplo de la demanda por celda mientras que la figura 6 presenta el resultado de una infraestructura eléctrica.

Modelo Vectorial: La principal diferencia entre este modelo y el anterior consiste en que en este modelo no se crea una cuadrícula, sino que toda la simulación se realiza de forma vectorial. De esta forma, podemos re-utilizar tanto la localización de las áreas urbanizables y de las infraestructuras relevantes como la densidad de suministros al ser estos también valores que podemos calcular vectorialmente. Por otro lado, para repartir los nuevos suministros usaremos un modelo por agentes de software [40]. Cada agente modela un nuevo suministro que se quiere establecer de forma que tratará de localizar el punto que maximice sus gustos. Estos gustos se generan a partir de los resultados de la figura 4.

Cada agente poseerá un tipo de inteligencia distinta así como una capacidad de cálculo diferente que se traducirá en la cantidad de preguntas que podrá hacer en el entorno para descubrir si los puntos donde se puede asentar cumplen con sus gustos o no. Nótese que los agentes compiten por asentarse en las zonas libres de distintas áreas urbanizables.

Para finalizar, cada área urbanizable es asignada a la infraestructura eléctrica mediante un Diagrama de Voronoi [2].

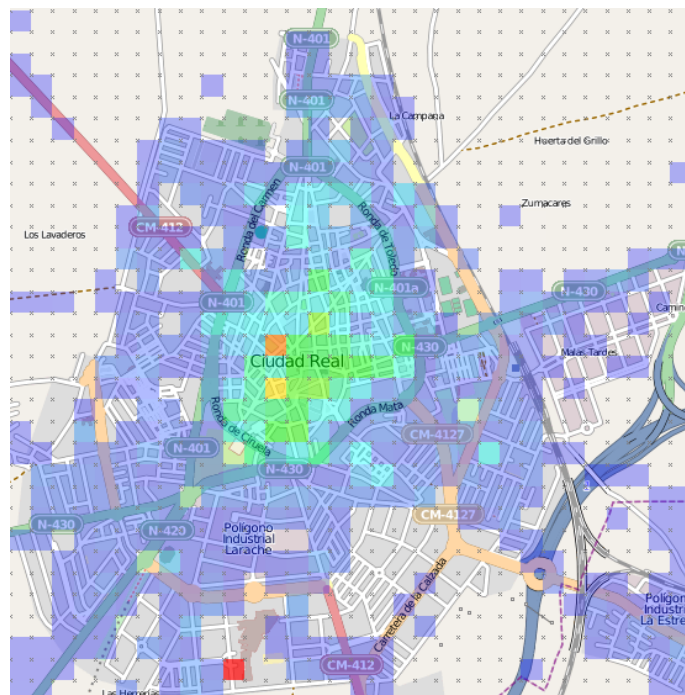


Figura 5: Potencia demandada.

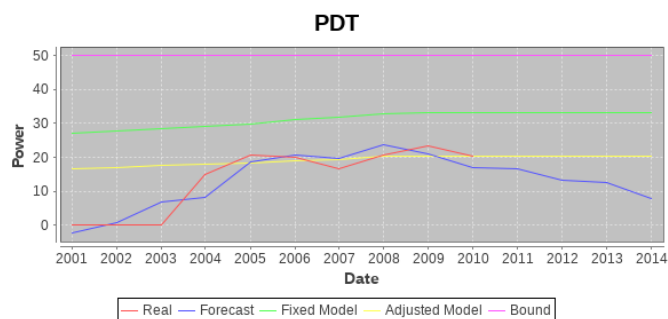


Figura 6: Datos reales y predicción para una subestación.

2.2. Inspección de líneas eléctricas mediante vehículos aéreos no tripulados

En este proyecto se pretende realizar una inspección integral de las líneas eléctricas aéreas mediante Vehículos Aéreos no Tripulados (UAV). Para ello, el UAV recorre las líneas eléctricas aéreas tomando imágenes de estas y de sus alrededores. Mediante el procesamiento de estas imágenes es posible realizar la inspección y determinar las distancias mínimas de seguridad exigidas por la legislación entre las líneas eléctricas y los elementos de su entorno, principalmente con la vegetación y/o edificaciones [19].

Las imágenes obtenidas durante los vuelos de inspección serán procesadas por varios algoritmos de visión artificial que serán los encargados de emitir los informes sobre el estado de las instalaciones. No obstante, para reducir los tiempos de procesamiento de las imágenes y optimizar el sistema de inspección es necesario desarrollar un GIS que proporcione información sobre el entorno de las líneas eléctricas al sistema de inspección [18] con el fin de determinar el algoritmo óptimo de procesamiento en cada situación. Para ello, el sistema de inspección deberá contar con información relevante sobre los elementos clave para la inspección de las líneas eléctricas definidos en la legislación. Se obtendrán las coordenadas y distancias características de estos elementos respecto a las líneas eléctricas. Los elementos a inspeccionar son:

- Áreas edificadas
 - Edificios particulares
 - Pueblos y núcleos urbanos
 - Campos de fútbol
 - Piscinas
 - Obras
- Zonas de posible peligro para el UAV
- Carreteras y calles
- Líneas del ferrocarril
- Ríos
- Zonas de arbolado y árboles aislados
- Postes y torres eléctricas

Parte de la información que se necesita ya estaba disponible en los datos provistos por OSM. No obstante, estos datos no tenían en muchos casos la precisión ni el detalle requerido y en otros no existía información de ningún tipo sobre los elementos implicados en la inspección de las líneas eléctricas aéreas. Por ello, los datos necesarios se han obtenido de GeoEuskadi. Estos datos han sido transformados al formato de datos OSM y a continuación incluidos en una bases de datos PostgreSQL [13] con la extensión PostGIS [34] desde la cual se puede acceder fácilmente a ellos.

El proceso de importación ha consistido en la traducción de las geometrías y de los metadatos usando el software ogr2osm [12], osm2pgsql [27] y adaptando las reglas de traducción en base a la documentación de GeoEuskadi para realizar la conversión de los ficheros originales en formato Shapefile (SHP) al formato OSM que nos interesa.

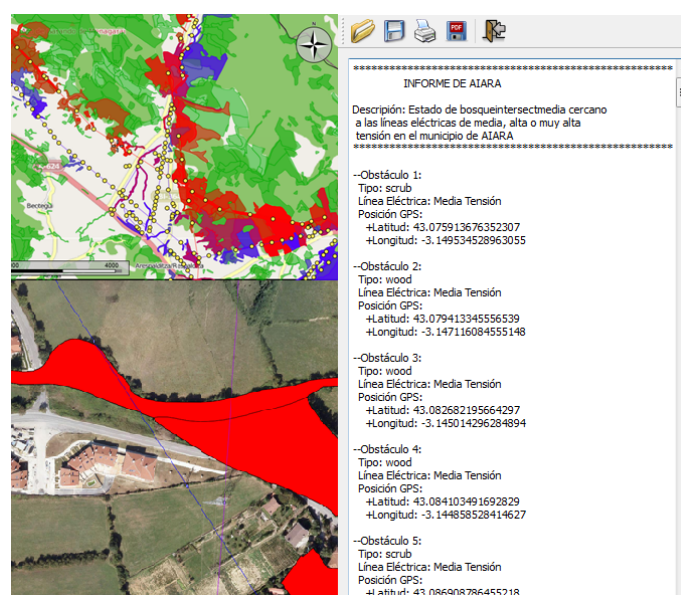


Figura 7: Cercanía de la vegetación a las líneas eléctricas.

La extensión PostGIS facilita en gran medida la ejecución de las consultas espaciales requeridas. Aunque en los datos se disponga de toda la información cartográfica, en este caso de la Comunidad Autónoma Vasca, para la inspección de las líneas eléctricas aéreas únicamente es interesante conocer el entorno de las líneas eléctricas. Este filtrado de información es realizado mediante consultas a la base de datos PostgreSQL.

Una vez filtrada toda la información geográfica, se elaboran una serie de informes que son empleados por el sistema de inspección automático para determinar la presencia de determinados elementos en las imágenes antes de que estas sean procesadas. Gracias a esta información previa, las imágenes son procesadas únicamente con los algoritmos específicos necesarios para la detección y cálculo de las distancias a los elementos presentes en cada imagen. Existirán, por ejemplo, algoritmos específicos para la detección y cálculo de las distancias de la vegetación que trabajarán en las zonas características de esta vegetación dentro del espectro electromagnético. En cambio, para la detección de otros elementos la información utilizada estará centrada en otras regiones del espectro, siendo clave a la hora de seleccionar estas bandas la información geográfica proporcionada por el sistema GIS desarrollado.

En la figura 7 se puede observar a modo de ejemplo la vegetación presente a una distancia determinada de las líneas eléctricas aéreas de media tensión (morado) y de las de muy alta tensión (rojo), así como un informe automático con las zonas de vegetación cercanas a las líneas eléctricas aéreas de media, alta o muy alta tensión que deben ser tenidas en cuenta durante la inspección.

Como se puede apreciar en la zona superior izquierda de la figura 7, para el caso de la vegetación, a los datos ya disponibles en la base de datos de OSM se ha añadido para este proyecto toda la información sobre las líneas eléctricas aéreas (postes, torres,

tendido, etc), así como un mayor detalle de la vegetación presente en este caso en el municipio alavés de Ayala/Aiara que es donde se han realizado las pruebas de campo para la inspección de las líneas eléctricas [16].

2.3. Indicadores de sostenibilidad

Los indicadores se pueden definir como medidas en el tiempo de las variables de un sistema que nos dan información sobre las tendencias de éste, aspectos concretos que nos interesa analizar. Éstos pueden estar compuestos simplemente por una variable (número de vehículos de un municipio) o por un grupo de ellas, como por ejemplo los metros cuadrados de verde urbano por habitante o también pueden encontrarse interrelacionadas formando índices complejos, como los índices económicos. En la gestión ambiental se utilizan para tres propósitos:

- Suministrar información sintética con el fin de evaluar los problemas.
- Establecer objetivos.
- Controlar el cumplimiento de dichos objetivos.

Actualmente muchos de los datos necesarios para el cálculo de los indicadores de sostenibilidad provienen de estudios realizados por entidades públicas o privadas. Estos estudios suelen ser largos y tediosos, y los resultados finales no terminan siendo públicos o tardan demasiado tiempo en estar disponibles para los ciudadanos o la sociedad en general. Además, dado que los datos son obtenidos por diferentes entidades, estos indicadores se suelen presentar cada uno por separado sin estudiar las posibles relaciones entre ellos. Por tanto, la centralización y homogeneización de toda la información relacionada a estos indicadores permitirá facilitar enormemente esta labor.

En el presente proyecto se pretende solventar la falta de coordinación y eficiencia en el análisis de los indicadores de sostenibilidad desarrollando una plataforma que, incluyendo toda la información sobre los indicadores de sostenibilidad y medioambientales, facilite la toma de decisiones en estos ámbitos y presente de forma conjunta y en tiempo real toda la información relacionada disponible.

El Programa Bizkaia 21 (2011-2016) [3], se estructura en torno a los 10 compromisos de Aalborg+10 [35], identificando 34 líneas estratégicas y 93 objetivos concretos, que a su vez proponen 332 actuaciones, que deberán alcanzarse contando con la participación activa de todos los departamentos forales, así como con la acción municipal. Los criterios de sostenibilidad que la sustentan comprenden consideraciones de tipo social, ambiental, económico y cultural.

El objetivo de este proyecto es múltiple. Por un lado se pretende integrar, tanto como se pueda, el cálculo de indicadores medioambientales. Por otro lado se pretende mejorar de diversas maneras el cálculo, realizando análisis posteriores o calculando con mayor nivel de detalle algunos indicadores. Para ello hemos realizado una prueba piloto sobre el municipio de Sopelana en donde calculamos los siguientes indicadores medioambientales:

Indicador 2: Población con acceso a pie a áreas verdes.

Indicador 3: Calidad del aire ambiental local.

Indicador 6: Población expuesta a niveles de ruido excesivo.

Indicador 7: Uso sostenible del suelo.

Indicador 13: Ocupación del suelo urbano.

Indicador 19: Porcentaje de viviendas secundarias o desocupadas.

Indicador 23: Municipios con plan de accesibilidad.

Indicador 25: Distancia del municipio a centros sanitarios.

Indicador 30: Crecimiento urbano previsto.

La elección de Sopelana ha sido debida al ser un municipio pequeño, cercano y conocido.

Al igual que en los anteriores casos, los datos de OSM no eran todo lo completos que se deseaban. En particular aunque las vías urbanas e interurbanas están perfectamente cubiertas faltan sin embargo información referente a las construcciones urbanas, las zonas urbanizadas, localización de servicios básicos, zonas de esparcimiento y usos de suelo. Ésta información ha sido obtenidas a partir de los siguientes ficheros que se encuentran en GeoEuskadi:

- BTA_EDI_MANZANAS_A_5000_ETRS89
- BTA_EDI_ENTI_POBLACION_A_5000_ETRS89
- BTA_CUBIERT_TERRESTRE_A_5000_ETRS89
- BTA_EDI_EDIFICACIONES_A_5000_ETRS89
- BTA_SER_INSTALACIONES_A_5000_ETRS89
- CB_MUNICIPIOS_5000_ETRS89

El proceso de importación ha consistido, igual que en el anterior caso, en la traducción de las geometrías y de los metadatos usando el software ogr2osm y osm2pgsql. En este caso también se ha procedido a la inspección manual de los resultados en la que se han realizado un postproceso consistente en la simplificación de geometrías, eliminación de duplicidades, corrección manual de errores y completado de información a partir del conocimiento de la zona y las imágenes de satélites PNOA [26].

Posteriormente pasamos a definir distintas metodologías de cálculo para mejorar las estimaciones de los indicadores de sostenibilidad anteriormente citados. En [6] se puede visualizar una demostración de alguna de las metodologías propuestas.

2.3.1. Indicador 2: Población con acceso a pie a servicios básicos y áreas verdes

En la actualidad, para estimar este indicador se expande en 300 metros la geometría del servicio básico o área verde. Seguidamente, se comprueba qué núcleos poblacionales tienen intersección con este área, de forma que se define la población con acceso a pie a los servicios básicos y áreas verdes como la suma de los habitantes de dichas poblaciones. Se entiende aquí *núcleos poblacionales* como los entes más pequeños a los que se tenga acceso, que en este momento corresponden a poblaciones.

Esta estimación es claramente optimista por dos razones: se considera que todos los habitantes de una población están a la misma distancia del área verde o servicio básico y no se tiene en cuenta los accidentes geográficos limitantes como montañas o ríos que pueden dar lugar a amplios rodeos.

Con los datos provistos por la base de datos de OSM podemos mejorar esta estimación, pudiendo incluso llegar a calcular el valor exacto, si se tiene acceso a la información del padrón municipal. Nuestra metodología consiste en:

- Para cada área verde expandir 300 metros su geometría.
- Obtenemos todos los edificios que están dentro de dicha área.
- Obtenemos la población del indicador
 - Haciendo una estimación del número de habitantes por edificio tenemos una aproximación para la población con acceso a pie a áreas verdes.
 - En el caso de tener acceso al padrón municipal, se podría usar esa información para calcular este valor de forma exacta.

Esta metodología se puede mejorar aún más si en vez de expandir la geometría del área verde o infraestructura básica usamos un modelo de grafo de la localidad a partir de la información de vías. En este caso el primer punto de la metodología quedaría de la siguiente manera:

- Para cada área verde o infraestructura básica localizamos las vías susceptibles de ser usadas a pie que empiezan, terminan o pasan a través de ella.
- Seguimos dichas vías durante 300 metros teniendo en cuenta las posibles ramificaciones de estas vías (intersecciones con otras vías, puentes, pasos a nivel, etc).
- Localizamos todos los edificios que tienen acceso a dichas vías.
- Hacemos el paso 3 de la anterior metodología, es decir, haciendo una estimación del número de habitantes por edificio obtenemos una aproximación para la población con acceso a pie a áreas verdes. En el caso de tener acceso al padrón municipal, se podría usar esa información para calcular este valor de forma exacta.

2.3.2. Indicador 3: Calidad del aire ambiental local

Para el caso de la calidad del aire, se ha extendido la prueba sobre todo el área del Gran Bilbao. Actualmente la información disponible sobre la calidad del aire es un índice global para toda la comarca del Gran Bilbao y los índices parciales para cada estación [21].

Nuestra propuesta consiste en aumentar la resolución y la cantidad de información a nivel local proporcionando información sobre el índice de calidad del aire y la concentración media de algunos de los contaminantes implicados en su cálculo, logrando un detalle a nivel de manzana.

Para esta prueba piloto se ha empleado la información proporcionada por las estaciones de la Red de Vigilancia y Control de la Calidad del Aire disponible en nuestra

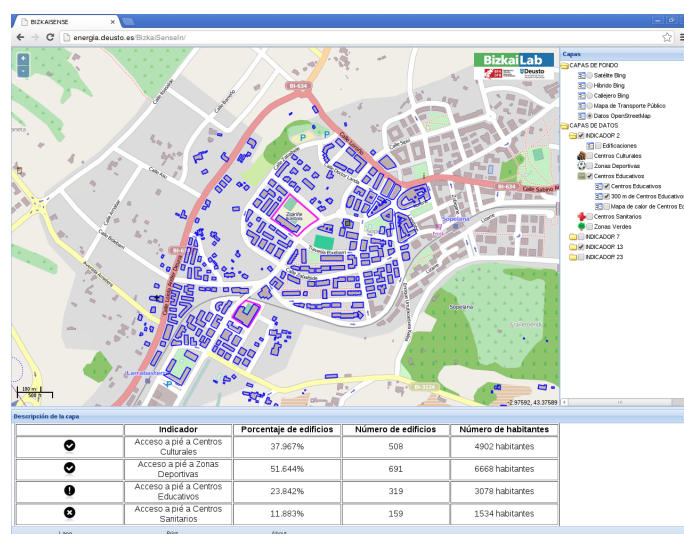


Figura 8: Edificaciones con acceso a pie a centros de educación.

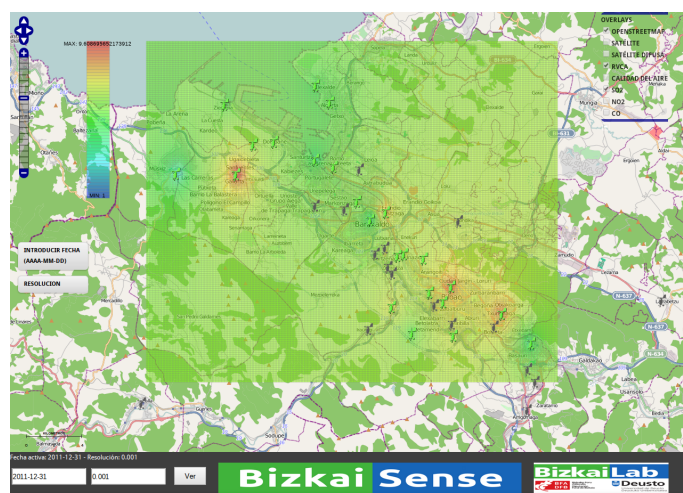


Figura 9: Edificaciones con acceso a pie a centros de educación.

base de datos [7], y empleando la metodología propuesta por el departamento de Medio Ambiente del Gobierno Vasco [22] se han calculado los índices parciales y global de la calidad del aire, así como la concentración media diaria de determinados contaminantes. A partir de esta información, se ha calculado la calidad del aire en cada punto dentro de todo el área del Gran Bilbao mediante interpolación por *Inverse Distance Weighting* de los datos proporcionados por cada estación de medición de la calidad del aire. La figura 9 muestra un ejemplo en el que se visualiza la concentración media de NO_2 en una fecha determinada [6].

No obstante, esta metodología es muy simple y se puede mejorar mediante el uso de un modelo de dispersión de contaminación que proporcione valores más fiables en función de la dirección y fuerza del viento así como de los obstáculos y la orografía presente en el territorio. De esta forma se pueden conseguir mapas de personas en riesgo de niveles

excesivos de exposición a contaminantes y/o analizar el número de personas que pueden estar expuestas a ellos.

2.3.3. Indicador 6: Población expuesta a niveles de ruido excesivo

El proceso de cálculo actual no está claramente definido en las memorias de los indicadores aunque se nombra el uso de mapas de ruido. Teniendo acceso a estos mapas de ruido (en un formato digital tipo Shapefile o equivalente) y al padrón municipal, se podría calcular exactamente la población expuesta a niveles de ruido. Nuestra metodología consistiría en:

- Vectorizar el mapa de ruido extrayendo las fronteras de los umbrales de cada categoría.
- Obtenemos todos los edificios que están dentro de cada área.
- Haciendo una estimación del número de habitantes por edificio obtenemos una aproximación para la población expuesta a cada nivel de ruido (incluso teniendo en cuenta las alturas de los edificios). En el caso de tener acceso al padrón municipal, se podría usar esa información para calcular este valor de forma exacta.

2.3.4. Indicadores 7, 13 y 19: Uso sostenible del suelo, ocupación del suelo urbano y porcentaje de viviendas secundarias o desocupadas

Sobre estos indicadores no se presentan mejoras, simplemente se muestra que también es posible calcularlos con esta herramienta. Hay que hacer notar, que en para el caso del indicador 19 los datos contenidos en GeoData Euskadi no son muy fiables al estar siempre vacíos los campos referentes al uso de la vivienda.

2.3.5. Indicador 23: Municipios con plan de accesibilidad

En la actualidad este indicador sólo constata si se tiene o no un plan de accesibilidad, no su cumplimiento efectivo ni su alcance. OSM cuenta con (al menos) datos referentes a la accesibilidad mediante discapacidades motoras [32] aunque es fácilmente ampliable para contener otros tipos de discapacidades tales como: a) indicadores visuales para sordos, b) placas en braille y señales sonoras para ciegos, c) texto adaptado para personas mayores o con discapacidades intelectuales, etc. Para cada infraestructura se guarda información sobre el grado de accesibilidad (en tres niveles: completamente accesible, con dificultad o imposible) por lo que es muy sencillo calcular un índice de accesibilidad a infraestructuras simplemente calculando el porcentaje de infraestructuras de cada tipo de accesibilidad.

2.3.6. Indicador 25: Distancia del municipio a centros sanitarios

El proceso de cálculo actual no está claramente definido en las memorias de los indicadores y existen varias formas de calcularlo: en línea recta o siguiendo las vías. En el primer caso se cometen grandes errores, como ya se ha comentado anteriormente, al no tenerse en cuenta los accidentes geográficos, como montañas o ríos, mientras que en el segundo caso no se tiene en cuenta las características de las vías que pueden hacer que distancias muy cortas lleven un tiempo excesivo (o incluso imposible) en caso de incidencias meteorológicas. La metodología que se propone en este proyecto para mejorar el cálculo de este indicador es:

- Calcular el tiempo de viaje desde cada edificio al centro sanitario más próximo respetando los límites de velocidad que tienen las diferentes vías. Realizar el cálculo suponiendo distintos escenarios meteorológicos.
- De nuevo, estimando el número de personas que hay en cada edificio, se puede conseguir la población que tiene acceso a un centro sanitario en cada franja de tiempo y, con acceso a los datos del padrón municipal, el valor exacto.

2.3.7. Indicador 30: Crecimiento urbano previsto

A pesar del nombre, este indicador sólo presenta el crecimiento urbano que se ha producido en los distintos grupos de edades entre 2001 y el presente año sin realizarse una estimación del número de habitantes en el futuro.

La metodología propuesta consiste en utilizar el modelo descrito en el primer caso para realizar una predicción a futuro no solo del crecimiento de la población sino de la dispersión urbana [4].

Nótese que de esta forma se podría calcular todos los índices anteriores en los distintos escenarios de crecimiento poblacional con el nivel de detalle propuesto. Además, al estar los modelos de crecimiento urbano basados tanto en las preferencias de los distintos tipos de habitantes como en las equipaciones que poseen los distintos municipios, será posible evaluar el impacto *a priori* tanto en los indicadores medioambientales, como en la variación poblacional que puede tener el desarrollo de ciertas infraestructuras y/o políticas.

2.4. Otros casos de éxito

Las posibles aplicaciones de datos geográficos voluntarios abarcan un gran rango de áreas. A continuación se enumeran otros casos en los que los datos geográficos voluntarios han jugado un papel muy importante.

2.4.1. Technical Cognitive Assistance Systems

Usando como soporte la aplicación geográfica Marble [20] de KDE [17] y datos de OSM, se están desarrollando varios proyectos de detección automática de señalética y sistemas avanzados de conducción. Estos proyectos buscan por un lado automatizar el reconocimiento de señales, comprensión y, en caso de confirmación por parte del usuario, su inclusión en los datos de OSM. Y por otro, utilizando estos datos y un conjunto de sensores, desarrollar sistemas conducción autónoma.

2.4.2. Election Results Maps

El diario El Mundo y Esri España desarrollaron un mapa interactivo [8] con motivo de las Elecciones Generales del 20 de Noviembre de 2011. En él se muestran sobre un mapa vivo todos los detalles de las votaciones y resultados de las elecciones, pudiendo alternar capas de resultados por comunidad o municipio. Además, permite la comparación de estos datos con resultados de elecciones anteriores y relacionarlos con otras capas de datos económicos y demográficos.

2.4.3. FlightGear Terrain Simulator

FlightGear [10] es un simulador de vuelo de código abierto construido sobre el motor de simulación SimGear. Este motor, con la ayuda de un *script*, permite generar escenarios con la red de carreteras y las edificaciones tomadas de OSM [11]. Por desgracia, para una generación más realista es recomendable que los edificios cuenten con los *tags* de descripción de altura y aproximadamente sólo un 2 % de los edificios especifican dicha información.

2.4.4. FireFIGHT

El proyecto FireFIGHT pretende el desarrollo de un sistema de optimización de la logística de tierra en incendios forestales. Está siendo desarrollado por la Universitat Politècnica de Catalunya. En él, los datos geográficos voluntarios juegan un papel doble. Por un lado, persiguen dotar a los simuladores de fuego de datos que puedan alterar el comportamiento de un fuego como ríos, carreteras, cultivos o distintos tipos de suelo y por otro, a la hora de trazar las rutas, contar con datos detallados y actualizados sobre las limitaciones de la red de carreteras de montaña para optimizar el despliegue de vehículos en caso de emergencias.

3. Conclusiones

La creciente disponibilidad de datos de información geográfica hacen posible el desarrollo de nuevas herramientas que acerquen la comunidad a los investigadores. De esta forma se produce una simbiosis altamente beneficiosa para ambas partes. Por un lado los investigadores cuentan con amplias bases de datos y herramientas para desarrollar innovadores proyectos de investigación. Por otro lado, la comunidad consigue apoyo, casos de uso y reconocimiento en general, cuando no colaboración a la hora de desarrollar o difundir su mensaje.

Agradecimientos

Este trabajo ha sido parcialmente subvencionado por los proyectos: ■ Cenit ENERGOS (CDTI-CEN2009-1048), ■ ITEA2 Nemo&Codec (CDTI-IDI-20110864-65), ■ Relifo (Iberdrola), ■ BizkaiSense (Programa BizkaiLab-DFB).

Referencias

- [1] Cruz E. Borges Ander Pijoan. *Wiki de Cat2Osm*. Url: <http://wiki.openstreetmap.org/wiki/Cat2Osm>.
- [2] Franz Aurenhammer. «Voronoi diagrams: a survey of a fundamental geometric data structure». En: *ACM Comput. Surv.* 23.3 (1991), págs. 345-405. ISSN: 0360-0300. DOI: 10.1145/116873.116880. Url: <http://doi.acm.org/10.1145/116873.116880>.
- [3] Diputación Foral de Bizkaia. *Bizkaia 21*. Url: <http://www.bizkaia21.net/>.
- [4] C.E. Borges, Y. Peña y A. Pijoan. «Agent Based Spatial Load Forecasting». En: *Proceedings of 3rd International Workshop on Agent Technologies for Energy Systems (ATES 2012) held at the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*. Valencia (Spain), 5th June 2012.
- [5] G.E.P. Box, G.M. Jenkins y G.C. Reinsel. *Time series analysis: forecasting and control*. Forecasting and Control Series. Prentice Hall, 1994. ISBN: 9780130607744.
- [6] Universidad de Deusto. *Unidad Energía del Instituto Tecnológico Deusto -- DeustoTech*. Url: <http://energia.deusto.es>.
- [7] DeustoTech. *BizkaiaSense*. Url: <http://helheim.deusto.es/bizkaisense>.
- [8] ESRI España. *Votos y Cifras, elecciones generales*. Url: http://www.esridemos.com/app/votosycifras_generales/.
- [9] Instituto Nacional de Estadística. *Instituto Nacional de Estadística*. Url: <http://www.ine.es>.

- [10] Desarrolladores y usuarios de FlightGear. *FlightGear*. Url: <http://flightgear.org/>.
- [11] Desarrolladores y usuarios de FlightGear. *FlightGear*. Url: http://wiki.flightgear.org/OpenStreetMap_buildings.
- [12] Open Source Geospatial Foundation. *ogr2ogr*. Url: <http://www.gdal.org/ogr2ogr.html>.
- [13] PostgreSQL Global Development Group. *PostgreSQL*. Url: <http://www.postgresql.org>.
- [14] Ministerio de Hacienda y Administraciones Públicas. *Sede electrónica de Catastro*. Url: <http://www.catastro.meh.es/>.
- [15] A. Hernando y col. «The workings of the maximum entropy principle in collective human behaviour». En: *Journal of The Royal Society Interface* 10.78 (2013). DOI: 10.1098/rsif.2012.0758. Url: <http://rsif.royalsocietypublishing.org/content/10/78/20120758.abstract>.
- [16] Iberdrola. *Pruebas de Iberdrola para la revisión de líneas eléctricas*. Url: <http://multimediamprofesionales.iberdrola.es/Noticias/IBERDROLA,REALIZA,PRIMERA,PRUEBA,HELICOPTERO,TRIPULADO,PARA,REVISION,LINEAS,ELECTRICAS,596.html>.
- [17] Proyecto KDE. *KDE*. Url: <http://kde.org>.
- [18] Juan I. Larrauri, Gorka Sorrosal y Mikel González. «A predictive model based on a cartographic and geodetic system for overhead power line inspection by UAV». En: *Proceedings of 8th International Conference on Intelligent Unmanned Systems (ICIUS 2012)*. 2012. ISBN: 978-981-07-4225-6.
- [19] Juan I. Larrauri, Gorka Sorrosal y Mikel González. «Automatic system for detection of obstacles and hotspots in the overhead powerline». En: *Proceedings of Remotely Piloted Aircraft Systems (RPAS 2012)*. 2012.
- [20] Contribuidores al proyecto Marble. *Marble*. Url: <http://edu.kde.org/marble/>.
- [21] Departamento de Medio Ambiente y Política Territorial del Gobierno Vasco. *Información de la calidad del aire en Euskadi*. Url: http://www.ingurumena.ejgv.euskadi.net/r49-n82/es/vima_ai_vigilancia/indice.apl?lenguaje=c.
- [22] Departamento de Medio Ambiente y Política Territorial del Gobierno Vasco. *Información de la calidad del aire en Euskadi*. Url: http://www.ingurumena.ejgv.euskadi.net/r49-3614/es/contenidos/informacion/red_calida_aire_capv/es_975/indice_calidad_c.html.
- [23] Agencia Estatal de Meteorología. *Agencia Estatal de Meteorología*. Url: <http://www.aemet.es>.
- [24] J. Moral Carcedo y J. Vicéns Otero. I.L.R Klein, 2003. Url: http://www.econolatin.com/klein/contenido/investigacion/pdf/DT_4_JMC_JV0.pdf.

- [25] J.M. De Diego Alarcón y A. Pérez-Navarro N. Fernández Terrones. «Comparativa entre OpenStreetMap y Cartociudad: caso de estudio de Valencia». En: *Actas de las VI Jornadas de SIG Libre, Girona 21--23 marzo, España*. Servicio de SIG y Teledetección - SIGTE - de la Universitat de Girona, 2012.
- [26] Instituto Geográfico Nacional. *PNOA*. url: <http://www.ign.es/PNOA>.
- [27] Colaboradores habituales de OpenStreetMap. *osm2pgsql*. url: <http://svn.openstreetmap.org/applications/utils/export/osm2pgsql/>.
- [28] Comunidad de usuarios de OpenStreetMap. *Herramientas de análisis de OpenStreetMap*. url: http://wiki.openstreetmap.org/wiki/List_of_OSM_based_Services.
- [29] Comunidad de usuarios de OpenStreetMap. *Herramientas de edición de OpenStreetMap*. url: <http://wiki.openstreetmap.org/wiki/Editing>.
- [30] Comunidad de usuarios de OpenStreetMap. *Herramientas de visualización de OpenStreetMap*. url: <http://wiki.openstreetmap.org/wiki/Category:Renderers>.
- [31] Comunidad de usuarios de OpenStreetMap. *Política de subida de datos de OpenStreetMap*. url: <http://www.openstreetmap.org/copyright>.
- [32] Comunidad de usuarios de OpenStreetMap. *Tags de accesibilidad de OpenStreetMap*. url: <http://wiki.openstreetmap.org/wiki/Accessibility>.
- [33] Comunidad de usuarios de OpenStreetMap. *Web OpenStreetMap*. url: <http://www.openstreetmap.org/>.
- [34] Refrations Research. *PostGIS*. url: <http://postgis.refrations.net>.
- [35] The Aalborg Commitments Secretariat. *Aalborg+10*. url: <http://www.aalborgplus10.dk/>.
- [36] David Stadelmann. *Which Factors Capitalize into House Prices? A Bayesian Averaging Approach*. CREMA Working Paper Series. Center for Research in Economics, Management y the Arts (CREMA), 2009. url: <http://EconPapers.repec.org/RePEc:cra:wpaper:2009-10>.
- [37] Heng Sun, Wayne Forsythe y Nigel Waters. «Modeling Urban Land Use Change and Urban Sprawl: Calgary, Alberta, Canada». English. En: *Networks and Spatial Economics* 7 (4 2007), págs. 353-376. issn: 1566-113X. doi: 10.1007/s11067-007-9030-y. url: <http://dx.doi.org/10.1007/s11067-007-9030-y>.
- [38] V. Vapnik. *Statistical Learning Theory*. John Willey & Sons, 1998.
- [39] Gobierno Vasco. *Infraestructura de Datos Espaciales - IDE de Euskadi*. url: <http://www.geo.euskadi.net/s69-15375/es/>.
- [40] Michael Wooldridge y Nicholas R. Jennings. «Intelligent Agents: Theory and Practice». En: *Knowledge Engineering Review* 10 (1995), págs. 115-152.